

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 1, 2001, 16:16:58 ; Search time 91.75 Seconds
(without alignments)
21.119 Million cell updates/sec

Title: US-09-331-631a-21_COPY_32_91
Perfect score: 343
Sequence: 1 TENPCAQRCLQSCQEPDDL.....DTGATNORHPGERTGRGP 60

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	343	100.0	614	1 AH11_ARAHY	P43237 arachis hyp
2	318	92.7	626	1 AH12_ARAHY	P43238 arachis hyp
3	98.5	28.7	639	1 GLCA_SOYBN	P11827 glycine max
4	93	27.1	605	1 GLCA_SOYBN	P13916 glycine max
5	66.5	19.4	771	1 SM3A_HUMAN	O14563 homo sapien
6	66	19.2	605	1 VCLA_GOSHI	P09799 gossypium h
7	64	18.7	1300	1 IRR_CAVPO	P14617 cavia porce
8	63.5	18.5	588	1 VCLA_GOSHI	P09801 gossypium h
9	63.5	18.5	772	1 SM3A_MOUSE	O08685 mus musculu
10	63.5	18.5	772	1 SM3A_MOUSE	O08685 mus musculu
11	63	18.4	591	1 UL49_EBV	P14347 Epstein-bar
12	61	17.8	545	1 ACLA_SOLTU	P31684 solanum tub
13	61	17.8	545	1 ACLA_SOLTU	P31685 solanum tub
14	60.5	17.6	777	1 SM3D_HUMAN	O95025 homo sapien
15	60.5	17.6	4393	1 PGBM_HUMAN	P98160 homo sapien
16	60	17.5	419	1 VEGC_HUMAN	P49767 homo sapien
17	59.5	17.3	772	1 SM3A_CHICK	O90607 gallus gall
18	59	17.2	147	1 IP2K_SOLTU	O41435 solanum tub
19	59	17.2	356	1 TRBM_BOVIN	P06579 bos taurus
20	59	17.2	448	1 FBLS_MOUSE	O9WV89 mus musculu
21	59	17.2	448	1 FBLS_MOUSE	O9WV89 mus musculu
22	58.5	17.2	751	1 SM3C_CHICK	O42236 gallus gall
23	58.5	17.1	106	1 THG1_NICPA	O24115 nicotiana p
24	58.5	17.1	778	1 SZ1B_BRARE	O9W686 brachyranio
25	58.5	17.1	785	1 SM3E_CHICK	O42237 gallus gall
26	58.5	17.1	4753	1 LRP_CAEEL	O04833 caenorhabdi
27	58	16.9	448	1 FBLS_HUMAN	O9WV89 mus musculu
28	58	16.9	487	1 VSWB_TRYBR	P20947 trypanosoma
29	57.5	16.8	707	1 BMPI_XENLA	P98070 xenopus lae
30	57.5	16.8	2768	1 THYG_HUMAN	P01266 homo sapien
31	57	16.6	88	1 OM3_CHI2R	P21355 chlamydia t
32	57	16.6	153	1 IP2K_SOLTU	P01080 solanum tub
33	57	16.6	751	1 SM3C_HUMAN	O99985 homo sapien

34	57	16.6	751	1 SM3C_MOUSE	O62181 mus musculu
35	57	16.6	823	1 CN8A_MOUSE	O88502 mus musculu
36	56	16.3	147	1 IP2K_SOLTU	O41489 solanum tub
37	56	16.3	1382	1 INSR_HUMAN	P06213 homo sapien
38	56	16.3	2813	1 VWF_CANFA	O28295 canis famli
39	56	16.3	4344	1 LRP1_HUMAN	O07934 homo sapien
40	55.5	16.2	562	1 ATP2_HEYBR	P29685 hevea bras
41	55.5	16.2	730	1 BMPI_HUMAN	P13497 homo sapien
42	55.5	16.2	761	1 SM3D_CHICK	O90663 gallus gall
43	55.5	16.2	991	1 BMPI_MOUSE	P98063 mus musculu
44	55.5	16.2	1106	1 STC_DROME	P40798 drosophila
45	53.5	16.2	2476	1 ZAN_PIG	O28983 sus scrofa

ALIGNMENTS

RESULT	1																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																				
--------	---	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

```

OC Arachis hypogaea (Peanut) .
OC Eukaryota: Viridiplantae; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
OC Fabales; Fabaceae; Papilionoideae; Arachis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. FLORUNNER.
RX MEDLINE=96013631; PubMed=7560062;
RT Burks A.W., Cockrell G., Stanley J.S., Helm R.M., Bannon G.A.;
RT "Recombinant peanut allergen Ara h I expression and IgE binding in
RT patients with peanut hypersensitivity."
RT J. Clin. Invest. 96:1715-1721(1995) .
CC -I- SIMILARITY: TO 7S SEED STORAGE PROTEINS (PHASOLIN, VICILIN,
CC CONVICILIN, CONGLYCININ, ETC.) .
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sdb.ch/announce/
CC or send an email to license@isb-sdb.ch) .
CC -----
CC EMBL: L34402; AAB00861.1; -.
CC HSSP: P50477; ICAM.
DR INTERPRO: IPR001113; -.
DR PfAM: PF00546; Seedstore_7s; 1.
RW Allergen.
SO SEQUENCE 626 AA; 71345 MW; 1A6BBBE4149D0E3 CRC64;

Query Match 92.7%; Score 318; DB 1; Length 626;
Best Local Similarity 90.6%; Pred. No. 3.9e-28;
Matches 58; Conservative 0; Mismatches 2; Indels 4; Gaps 1;

Oy 1 TENPACORCLGCGOQEPDDLKQACESRCTKLEYPCRYD---TGATNORHPGERTR 56
Db 34 TENPACORCLGCGOQEPDDLKQACESRCTKLEYPCRYDPRGHTGTTNORSPGERTR 93
Oy 57 GRQP 60
Db 94 GRQP 97

RESULT 3
ID GLCX__SOYBN STANDARD: PRT; 639 AA.
AC P11827;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 01-AUG-1992 (Rel. 23, Last annotation update)
DE BETA'-CONGLYCININ, ALPHA' CHAIN PRECURSOR.
GN CG-1.
OS Glycine max (Soybean) .
OC Eukaryota: Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
OC Fabales; Fabaceae; Papilionoideae; Glycine.
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=86250867; PubMed=3013879;
RA Doyle J.J., Schuler M.A., Godette W.D., Zenger V., Beachy R.N.,
RA Slightom J.L.;
RT "The glycosylated seed storage proteins of Glycine max and Phaseolus
RT vulgaris. Structural homologies of genes and proteins."
RT J. Biol. Chem. 261:9228-9238(1986) .
RN [2]
RP SEQUENCE OF 340-639 FROM N.A.
RC MEDLINE=83143289; PubMed=6897678;
RA Schuler M.A., Ladin B.F., Pollaco J.C., Freyer G., Beachy R.N.;
RT "Structural sequences are conserved in the genes coding for the
RT alpha, alpha' and beta-subunits of the soybean 7S seed storage
RT protein."
RL Nucleic Acids Res. 10:8245-8261(1982) .

```

```
CC -I- FUNCTION: SEED STORAGE PROTEIN. ACCUMULATES DURING SEED
CC DEVELOPMENT AND IS HYDROLYZED AFTER GERMINATION TO PROVIDE A
CC CARBON AND NITROGEN SOURCE FOR THE DEVELOPING SEEDLING.
CC -I- SUBUNIT: THE ALPHA-, ALPHA-, AND BETA-SUBUNITS ASSOCIATE IN
CC VARIOUS COMBINATIONS TO FORM TRIMERIC PROTEINS.
CC -I- SUBCELLULAR LOCATION: CYTOLEDMARY MEMBRANE-BOUND VACUOLAR PROTEIN
CC BOOIES.
CC CC -I- SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN,
CC CONVICTILIN, CONGLYCININ, ETC.).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC The European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M13759; AAB01374.1; -.
DR EMBL; J01290; .. NOT_ANNOTATED_CDS.
DR PIR; B24810; B24810.
DR HSSP; P02853; 2PHL.
DR INTERPRO; IPR001113; .
DR PFAM; PF00546; Seedstore_7s; 2.
KW Seed storage protein; Signal; Glycoprotein; Multigene family.
FT SIGNAL 1 22
FT PROPEP 23 62
FT CHAIN 63 639 BETA-CONGLYCININ, ALPHA' CHAIN.
FT CARBOHYD 277 277 N-LINKED (GLCNAG. . .) (POTENTIAL).
FT CARBOHYD 551 551 N-LINKED (GLCNAG. . .) (POTENTIAL).
FT CONFIDENT 543 543 P -> L (IN REF. 2).
FT CONFLICT 549 549 M -> V (IN REF. 2).
FT CONFLICT 608 608 S -> T (IN REF. 2).
SQ SEQUENCE 639 AA; 74325 MW; 469BF24C79651E3F CRC64;

Query Match      28.7%; Score 98.5; DB 1; Length 639;
Best Local Similarity 30.6%, Pred. No. 0.00083;
Matches 22; Conservative 10; Mismatches 25; Indels 15; Gaps 3;

QY 2 ENPCAPORCTGSCQQPPDLKOKACECSRCT--KLEVDPRCVDTGT-----ATNQR 48
   ||| | | | | : | | | | | | | : | | | | | : | | | | | | | | |
DB 31 QNPSPNKCLRSNSKSDSYRNQACHARNLKVEEEEC--EGQLIPRPHPEREROO 88
QY    49 HPGERTRGRQP 60
     | | | | |
DB    89 HGKEDEGEQRP 100

RESULT 4
GLCA_SOYBN STANDARD; PRT; 605 AA.
ID AC PI3916;
PT 01-JAN-1990 (Rel. 13, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE BETA-CONGLYCININ, ALPHA CHAIN PRECURSOR.
OS Glycine max (Soybean).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
OC Fabales; fabaceae; Papilionoideae; Glycine.
NM [ ]
NP SEQUENCE FROM N.A.
RC TISSUE=COYLEDON;
RX MEDLINE=91355860; PubMed=2103438;
RA Sebastiani F.L., Farrel L.B., Schuler M.A., Beechy R.N.;
RT "Complete sequence of a cDNA of alpha subunit of soybean beta-
conglycinin."
RL Plant Mol. Biol. 15:197-201(1990).
CC -I- FUNCTION: SEED STORAGE PROTEIN. ACCUMULATES DURING SEED
CC DEVELOPMENT AND IS HYDROLYZED AFTER GERMINATION TO PROVIDE A
CC CARBON AND NITROGEN SOURCE FOR THE DEVELOPING SEEDLING.
CC -I- SUBUNIT: THE ALPHA-, ALPHA-, AND BETA-SUBUNITS ASSOCIATE IN
```

```

CC      VARIOUS COMBINATIONS TO FORM TRIMERIC PROTEINS.
CC      -I- SUBCELLULAR LOCATION: EMBRYO AXIS, AND COTYLEDONARY MEMBRANE-BOUND
CC      VACUOLAR PROTEIN BODIES.
CC      -I- SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN,
CC      CONVICILIN, CONGLYCININ, ETC.).
CC      -----
CC      THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; X17698; CAA35691.1; -.
DR      PIR; S14681; FMSYBA.
DR      HSSP; P50477; ICAM.
DR      INTERPRO; IPR001113; -.
DR      Pfam; PF00546; Seedstore_7s; 1.
KW      Seed storage protein; Signal; Glycoprotein; Multigene family.
FT      SIGNAL 1 22
FT      PROPEP 23 62
FT      CHAIN 63 605
FT      CARBOHYD 261 261
FT      CARBOHYD 517 517
SO      SEQUENCE 605 AA; 70293 MW; CBBDA30506BBBC57 CRC64;

Query Match          27.1%; Score 93; DB 1; Length 605;
Best Local Similarity 31.4%; Pred. NO. 0.0032;
Matches 22; Conservative 8; Mismatches 16; Indels 24; Gaps 3;

QY      2 ENPCAQRLGSCQGPDDLKQKACESCRTKLEYD-----PCRYDYGATNQRHP 50
      ||| :||| | | : || || | : |||
Db      31 ENPKHKKCTGSCNSRDSRYNQACHARCNLKVEKECEFEGETPRPR-----PRQHP 83
QY      51 -----PGER 54
      |||:
Db      84 EREPOQGEK 93

RESULT 5
SM3A_HUMAN STANDARD; PRT; 771 AA.
ID SM3A_HUMAN
AC Q14563;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE SEMAPHORIN 3A PRECURSOR (SEMAPHORIN III) (SEMA III).
GN SEMA3A.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
[1]
RN SEQUENCE FROM N.A.
RP TISSUE=FETAL BRAIN;
RX MEDLINE=94094332; PubMed=8269517;
RA Kolodkin A.L., Mathes D.J., Goodman C.S.;
RT "The semaphorin genes encode a family of transmembrane and secreted
RT growth cone guidance molecules.";
RL Cell 75:1389-1399(1993).
RN 121
RN SEQUENCE OF 1-37 FROM N.A.
RP Moessner J., Minx P., Hinds K., Stromwalt C.;
RL submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN 131
RN SEQUENCE OF 39-182 FROM N.A.
RP Rohling T., Tin-Wollam A.M., Duckels G.;
RL submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
CC -I- FUNCTION: INDUCES THE COLLAPSE AND PARALYSIS OF NEURONAL GROWTH
CC CONES. COULD SERVE AS A LIGAND THAT GUIDES SPECIFIC GROWTH CONES
CC BY A MOTILITY-INHIBITING MECHANISM. BINDS TO THE COMPLEX
CC NEURODILIN-1/PLEXIN-1 (BY SIMILARITY).

```

```

CC -1- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
CC -1- DOMAIN: STRONG BINDING TO NEUROPHILIN IS MEDIATED BY THE CARBOXY
CC THIRD OF THE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE SEMAPHORIN FAMILY.
CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L26081; AAA65938.1;
CC EMBL; AC004451; AAC06185.1;
CC EMBL; AC004848; AAC78622.1;
CC MIM: 603961;
CC INTERPRO: IPR001627;
CC INTERPRO: IPR003006;
CC DR PFAM: PF01403; Sema; 1.
CC PFAM: PF00047; Ig; 1.
CC KW Signal: Immunoglobulin domain; Multigene family; Neurogenesis;
CC developmental protein; Glycoprotein.
CC SIGNAL 1 20
CC CHAIN 21 771
CC DOMAIN 240 538
CC FT DOMAIN 642 729
CC FT DOMAIN 727 769
CC FT DISULFD 649 722
CC FT CARBOHYD 53 53
CC FT CARBOHYD 125 125
CC FT CARBOHYD 590 590
CC SEQUENCE 771 AA; 88889 MW; 9985F8D3BAED8456 CRC64;
QY
QY Query Match 19.4%; Score 66.5; DB 1; Length 771;
QY Best Local Similarity 35.2%; Pred. No. 3.5;
QY Matches 19; Conservative 5; Mismatches 23; Indels 7; Gaps 2;
QY
QY 6 AQRCLSGSQGPEDDLKQKACESRCIKLEYDPRCYVDITATNQRHPGERTGRQ 59
QY 510 AQLPLHRC-----DIYGRACAECC--IARDPYCAMDGSGCSRYPFPTARRRRRQ 556
QY
RESULT 6
VCLA_GOSHI ID VCLA_GOSHI STANDARD; PRT; 605 AA.
AC P09799;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE VICILIN GC72-A PRECURSOR (ALPHA-GLOBULIN A).
OS Gossypium hirsutum (Upland cotton).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Malvales; Malvaceae; Gossypium.
OC [1]
RN SEQUENCE FROM N.A.
RA Chian C.A., Borroto K., Kamalay J.A., Dure L. III;
RT "Developmental biochemistry of cottonseed embryogenesis and
RT germination. XIX. Sequences and genomic organization of the alpha
RT globulin (vicilin) genes of cottonseed.";
RL Plant Mol. Biol. 9:533-546(1987).
CC -1- FUNCTION: SEED STORAGE PROTEIN.
CC -1- SUBCELLULAR LOCATION: CYTOLEDONARY MEMBRANE-BOUND VACUOLAR PROTEIN
CC BODIES.
CC -1- SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASOLIN, VICILIN,
CC CONVITICLIN, CONGLYCININ, ETC.).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

[illegible][illegible]

```

RT   the storage protein families";
RL   Plant Mol. Biol. 7:475-489(1986).
CC   -I- FUNCTION: SEED STORAGE PROTEIN.
CC   -I- SUBCELLULAR LOCATION: CYTOLEDMARY MEMBRANE-BOUND VACUOLAR PROTEIN BODIES.
CC   -I- SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASCOLIN, VICILIN, CONVULCILIN, CONVOLXICININ, ETC.).
-----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb.ch/announce/ or send an email to license@isb-sib.ch).
-----
DR   EMBL; M16891; AAA33071.1; -.
DR   PIR; A30838; FMCNAB.
DR   HSSP; P50477; ICAX.
DR   INTERPRO; IPR001113; -.
DR   DFRAM; PF00546; Seedstore_7s; 1.
KW   Seed storage protein; Signal.
FT   SIGNAL          1      25
FT   CHAIN           26     588
FT   FT              1      25
SO   SEQUENCE        588 AA; 69729 MW; 63E699B29A8BAD6B CRC64;
-----
Query Match               18.5%; Score 63.5; DB 1; Length 588;
Best Local Similarity    23.4%; Pred. No. 5.8;
Matches       18; Conservative 12; Mismatches       26; Indels 21; Gaps 3;

QY   5 CAORCLSQCGQEPPDLKQ-KACESRCIKLEYDP---RCVYGATGNQRHP----- 50
    | |||| :| | :: : | : | : | : | : | : | : |
DB   105 CQGRCLKRFEEQQGSGFOFCQHCHGDQGEPRKQGCVCRECRERYGENDNREREER 164
    | | | | | | | | | | | | | | | | | | | | | |

QY   51 -----FGERTGRKQP 60
    || : : |

DB   165 ABEETEGEGEQSHNP 181
-----
RESULT 9
SM3A_MOUSE STANDARD; PRT; 772 AA.
ID   SM3A_MOUSE
AC   O08665; Q62180; Q62215;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   SEMAPHORIN 3A PRECURSOR (SEMAPHORIN III) (SEMA III) (SEMAPHORIN D) (SEMA D).
DN   SEMA3A OR SEMAD OR SEMD.
GN   Mus musculus (Mouse).
OS   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. RN   [1]
RN   SEQUENCE FROM N.A.
RP   STRAIN=NMRI; TISSUE=EMBRYO;
RC   MEDLINE=95267431; PubMed=7748561;
RX   Puschel A.W., Adams R.H., Betz H.;
RA   "Murine semaphorin D/collapsin is a member of a diverse gene family RT and creates domains inhibitory for axonal extension.";
RL   Neuron 14:941-948(1995).
LN   [2]
RN   SEQUENCE FROM N.A.
RP   MEDLINE=97470885; PubMed=9331345;
RX   Taiguchi M., Ylase S., Fujisawa H., Naruse I., Saga S., Mishima M., RA   Yagi T.;
RT   "Distruption of semaphorin III/D gene causes severe abnormality in peripheral nerve projection.";
RL   Neuron 19:519-530(1997).
LN   [3]
RN   SEQUENCE OF 107-772 FROM N.A.
RP   TISSUE=FETAL BRAIN;
RC   MEDLINE=95267432; PubMed=7748562;
RX   MEDLINE=95267432; PubMed=7748562;

```

```

RA  Messersmith E.K., Leonardo E.D., Shatz C.J., Tessier-Lavigne M.,
RA  Goodman C.S., Kolodkin A.L.:
RA  "Semaforin III can function as a selective chemorepellent to pattern
RT  sensory projections in the spinal cord.";
RL  Neuron 14:949-959(1995).
CC  -I- FUNCTION: PLAY A ROLE IN GROWTH CONES GUIDANCE. MAY FUNCTION TO
CC  PATTERN SENSORY PROJECTIONS BY SELECTIVELY REPELLING AXONS THAT
CC  NORMALLY TERMINATE DORSALLY.
CC  -I- SUBCELLULAR LOCATION: SECRETED.
CC  -I- DEVELOPMENTAL STAGE: EXPRESSED EARLY IN EMBRYONIC DEVELOPMENT
CC  (E11) IN DISTINCT REGIONS OF THE NEUROECTODERM AND MESODERM.
CC  EXPRESSION BECAME MORE EXTENSIVE AT LATER STAGES.
CC  -I- DOMAIN: STRONG BINDING TO NEUROPILIN IS MEDIATED BY THE CARBOXY
CC  THIRD OF THE PROTEIN.
CC  -I- SIMILARITY: BELONGS TO THE SEMAPHORIN FAMILY.
CC  -I- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (see http://www.isb-sdb.ch/announce/
CC  or send an email to license@sdb.ch).
CC  -----
DR  EMBL; X85993; CAA59985.1; -.
DR  EMBL; D85028; BAA19773.1; -.
DR  EMBL; L40484; AAA73934.1; -.
DR  GDB; MGI:107558; SEMA3A.
DR  INTERPRO; IPR001627; -.
DR  INTERPRO; IPR003006; -.
DR  PFM; PF01403; Sema; 1.
DR  PFM; PF00047; Ig; 1.
KW  Signal; Immunoglobulin domain; Multigene family; Neurogenesis;
KW  Developmental protein; Glycoprotein.
FT  SIGNAL          1 20
FT CHAIN           21 772
FT DOMAIN          240 538
FT DOMAIN          643 730
FT DOMAIN          728 770
FT DISULFID        650 723
FT CARBOHYD        53 723
FT CARBOHYD        125 125
FT CARBOHYD        591 591
FT CONFLICT        193 193
FT CONFLICT        207 207
FT CONFLICT        253 253
FT CONFLICT        352 352
FT CONFLICT        403 403
FT CONFLICT        571 572
FT CONFLICT        616 620
FT CONFLICT        623 623
FT SEQUENCE        772 AA; 88799 MW; E89A08528B10AEC3 CRC64;
OY  6 AORCLOSCQOEPPDLKQACESCRTKLEYIPRCVYDTGATNQRRHPPGRRTRGRQ 59
OY  11 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB  510 AQLPLHRC-----DIYGRACAACC--LARDPYCAMDGSSGSRYPPTAKRRTRRQ 556
DB  18; Conservative 33.3%; Pred. No. 7.5; Mismatches 23; Indels 7; Gaps 2;

RESULT 10
SN3A_RAT
ID SM3A_RAT STANDARD; PRT; 772 AA.
AC 063548;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE SEMAPHORIN 3A PRECURSOR (SEMAPHORIN III) (SEMA III).
GN SEMA3A.

```

```

OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-WISTAR: TISSUE-BRAIN:
RX MEDLINE-97073089; Pubmed-8915837;
RA Giger R.J., Wolfer D.P., De Wit G.M.J., Verhaagen J.;
RT "Anatomy of rat semaphorin ii/collapsin-1 mRNA expression and
RT relationship to developing nerve tracts during neuroembryogenesis.";
RL J. Comp. Neurol. 375:378-392(1996).
CC -1- FUNCTION: MAY BE INVOLVED IN GUIDING GROWING AXONS TOWARDS THEIR
CC TARGETS BY FORMING A MOLECULAR BOUNDARY THAT INSTRUCTS AXONS TO
CC ENGAGE IN THE FORMATION OF SPECIFIC NERVE TRACTS. BINDS TO
CC NEUROFILIN.
CC -1- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
CC -1- DEVELOPMENTAL STAGE: AT E-11, EXPRESSION WAS RESTRICTED TO THE
CC OLFACTORY PIT, THE BASAL AND ROSTRAL SURFACE OF THE TELENCEPHALIC
CC VESICLE, THE EYE ANLAGE, THE EPITHELIUM OF RATHEE'S POUCH, AND
CC SOMITES. AT LATER DEVELOPMENTAL STAGES, IT WAS WIDELY DISTRIBUTED
CC IN NEURONAL AS WELL AS IN MESENCHYMAL AND EPITHELIAL STRUCTURES
CC OUTSIDE THE NERVOUS SYSTEM. AFTER BIRTH, MESENCHYMAL LEVELS
CC DECREASED RAPIDLY AND EXPRESSION BECAME RESTRICTED TO SPECIFIC
CC SETS OF NEURONS IN THE CNS. IN THE MATURE CNS, IT IS DETECTABLE IN
CC MITRAL CELLS, NEURONS OF THE ACCESSORY OLIV AND CEREBRAL CORTEX,
CC CEREBELLAR PURKINJE CELLS, AS WELL AS A SUBSET OF CRANIAL AND
CC SPINAL MOTONEURONS.
CC -1- DOMAIN: STRONG BINDING TO NEUROFILIN IS MEDIATED BY THE CARBOXY
CC THIRD OF THE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE SEMAPHORIN FAMILY.
CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: X95286; CAA64607.1; -.
DR INTERPRO: IPR001627; -.
DR INTERPRO: IPR003006; -.
DR PRAM: PF01403; Sema: 1.
DR PFAM: PF00047; Ig; 1.
KW Signal; Immunoglobulin domain; Multigene family; Neurogenesis;
KM Developmental protein; Glycoprotein.
FT SIGNAL 1 20
FT CHAIN 1 772
FT DOMAIN 21 772 SEMAPHORIN 3A.
FT DOMAIN 240 538 SEMA.
FT DOMAIN 643 730 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 728 770 ARG/LYS-RICH (BASIC).
FT DISULFID 650 723 BY SIMILARITY.
FT CARBOHYD 53 53 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 125 125 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 591 591 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 772 AA; 88808 MW; 240907812FF9F2D2 CRC64;

```

```

Query Match 18.5%; Score 63.5; DB 1; Length 772;
Best Local Similarity 33.3%; Pred. No. 7.5;
Matches 18; Conservative 6; Mismatches 23; Indels 7; Gaps 2;

```

```

QY 6 AGRCLSCQOEPPDLKOKACESEKCTKLEYDPKCVYDGTATNORHPGERTGRG 59
DB 510 AQLPLRHC-----DIYKACAEC--LARDPYCAMDSSCSKRYFTAKRRTRRQ 556

```

```

RESULT 11
UL49_EBV
ID UL49_EBV
AC P14347;
DT 01-JAN-1990 (Rel. 13, Created)
STANDARD; PRT; 591 AA.

```

```

DT 01-JAN-1990 (Rel. 13, Last sequence update)
DE 01-OCT-1996 (Rel. 34, Last annotation update)
DR PROTEIN BFRF2.
OS BFRF2.
GN Epstein-Barr virus (strain B95-8) (Human herpesvirus 4).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Lymphocryptovirus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-84270667; Pubmed-6087149;
RA Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,
RA Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Seguin C.,
RA Telford P.S., Watford B.G.;
RT "DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";
RL Nature 310:207-211(1984).
CC -1- SIMILARITY: BELONGS TO A FAMILY THAT GROUPS TOGETHER EBV BFRF2,
CC HVS-1 66, AND HCMV UL49.
CC -1- This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: V01555; CAA24880.1; ALT_INIT.
DR EMBL: V01555; CAA24880.1; ALT_INIT.
KM Early protein.
SQ SEQUENCE 591 AA; 63977 MW; EE6FF1E0721912E CRC64;

```

```

Query Match 18.4%; Score 63; DB 1; Length 591;
Best Local Similarity 34.4%; Pred. No. 6.6;
Matches 22; Conservative 5; Mismatches 19; Indels 18; Gaps 5;

```

```

QY 5 CAQRCLSCQOEPPDLKOKA-CESRCIKLEYDP--KCV-----YDTATNORHPG 53
DB 527 CQQR-----QAPEDARDEPLFSSCLEVELPQERCACGRLYQRYGTGPA--QAHPPGE 579
QY 54 RTRG 57
DB 580 AGCG 583

```

```

RESULT 12
ACIL_SOLTU
ID 4CL1_SOLTU
AC P31684;
STANDARD; PRT; 545 AA.
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE 4-COMMUNARATE--COA LIGASE 1 (EC 6.2.1.12) (4CL).
GN 4CL-1.
OS Solanum tuberosum (Potato).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I;
OC Solanales; Solanaceae; Solanum.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-91217100; Pubmed-2022667;
RA Becker-Andre M., Schulze-Lefert P., Hahlbrock K.;
RT "Structural comparison, modes of expression, and putative cis-acting
RT elements of the two 4-communarate: COA ligase genes in potato.*";
RL J. Biol. Chem. 266:8551-8559(1991).
CC -1- CATALYTIC ACTIVITY: ATP + 4-COMMUNARATE + COA = AMP + PYROPHOSPHATE
CC + 4-COMMUNARYL-COA.
CC -1- PATHWAY: BRANCH-POINT REACTIONS BETWEEN GENERAL PHENYLPROPANOID
CC METABOLISM AND PATHWAYS LEADING TO VARIOUS SPECIFIC END PRODUCTS.
CC -1- SIMILARITY: TO OTHER ENZYMES WHICH ACT VIA AN ATP-DEPENDENT
CC COVALENT BINDING OF AMP TO THEIR SUBSTRATE.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 RN [1]
 RP SEQUENCE FROM N A


```

FT DOMAIN 2536 2631 IG-LIKE C2-TYPE DOMAIN 11.
FT DOMAIN 2632 2728 IG-LIKE C2-TYPE DOMAIN 12.
FT DOMAIN 2729 2828 IG-LIKE C2-TYPE DOMAIN 13.
FT DOMAIN 2829 2926 IG-LIKE C2-TYPE DOMAIN 14.
FT DOMAIN 2927 3023 IG-LIKE C2-TYPE DOMAIN 15.
FT DOMAIN 3024 3114 IG-LIKE C2-TYPE DOMAIN 16.
FT DOMAIN 3115 3213 IG-LIKE C2-TYPE DOMAIN 17.
FT DOMAIN 3214 3300 IG-LIKE C2-TYPE DOMAIN 18.
FT DOMAIN 3301 3401 IG-LIKE C2-TYPE DOMAIN 19.
FT DOMAIN 3402 3490 IG-LIKE C2-TYPE DOMAIN 20.
FT DOMAIN 3491 3576 IG-LIKE C2-TYPE DOMAIN 21.
FT DOMAIN 3577 3671 IG-LIKE C2-TYPE DOMAIN 22.
FT DOMAIN 3671 3701 LAMININ G-LIKE 1 (GLOBULAR DOMAIN V A).
FT DOMAIN 3701 3847 EGF-LIKE 1.
FT DOMAIN 3846 3883 EGF-LIKE 2.
FT DOMAIN 3886 3924 LAMININ G-LIKE 2 (GLOBULAR DOMAIN V B).
FT DOMAIN 3924 3966 EGF-LIKE 3.
FT DOMAIN 3966 4104 EGF-LIKE 4.
FT DOMAIN 4106 4143 LAMININ G-LIKE 3 (GLOBULAR DOMAIN V C).
FT DOMAIN 4145 4178 HEPARAN SULFATE (POTENTIAL).
FT DOMAIN 4243 4391 HEPARAN SULFATE (POTENTIAL).
FT SITE 65 67 MEDIATES MOTOR NEURON ATTACHMENT
FT SITE 71 73 (POTENTIAL).
FT SITE 76 78 BY SIMILARITY.
FT SITE 4151 4153 BY SIMILARITY.
FT SITE 4301 4303 BY SIMILARITY.
FT DISULFID 199 212 BY SIMILARITY.
FT DISULFID 206 225 BY SIMILARITY.
FT DISULFID 219 234 BY SIMILARITY.
FT DISULFID 285 297 BY SIMILARITY.
FT DISULFID 292 310 BY SIMILARITY.
FT DISULFID 304 319 BY SIMILARITY.
FT DISULFID 325 337 BY SIMILARITY.
FT DISULFID 332 350 BY SIMILARITY.
FT DISULFID 344 359 BY SIMILARITY.
FT DISULFID 368 381 BY SIMILARITY.
FT DISULFID 375 394 BY SIMILARITY.
FT DISULFID 388 403 BY SIMILARITY.
FT DISULFID 766 775 BY SIMILARITY.
FT DISULFID 768 782 BY SIMILARITY.
FT DISULFID 785 794 BY SIMILARITY.

Query Match 17.68; Score 60.5; DB 1; Length 4393;
Best Local Similarity 32.28; Pred. No. 87;
Matches 19; Conservative 4; Mismatches 33; Indels 3; Caps 2;
OY 4 PCAQRCLQSCQEPDDLKQKACESRCTKLEDPKCVDTGATN--QRNPPGERTRGROP 60
Db 1137 PSCDDCDGTPTPTPSGLYGTCE-RCSCHGHSACPEPTGACGCGNHTESPCEGCGOP 1194

```

Search completed: March 1, 2001, 16:17:00
Job time: 440 sec

